

BIO392-cnv-freq_breast_carcinoma

Joelle Schneider

28.09.22

Step 1: Install package

```
if (!require(devtools)){
  install.packages('devtools')
}

## Loading required package: devtools
## Loading required package: usethis
if (!require(pgxRpi)){
  devtools::install_github('progenetix/pgxRpi')
}

## Loading required package: pgxRpi
library(pgxRpi)
```

Step2: breast carcinoma (C4872) NCIT code

Step3: Access the CNV frequency data from samples with breast carcinoma
codematches: T stands for true, takes only samples from the specific cancer!

```
freq <- pgxLoader(type='frequency', output='pgxseg', filters='NCIT:C4872',
                  codematches=T)
```

The retrieved data is an object containing two slots meta and data.

The meta slot looks like this:

```
freq$meta
```

##	code	label	sample_count
## 1	NCIT:C4872	Breast Carcinoma	15800
## 2	total		15800

The data slot includes two matrices. ## Here, the total and NCIT are the same. With more NCIT codes, the total would be a concatenated data slot

```
names(freq$data)
```

```
## [1] "NCIT:C4872" "total"
```

The frequency matrix looks like this

```
head(freq$data$`NCIT:C4872`)
```

```
##      filters reference_name  start    end gain_frequency loss_frequency no
## 1 NCIT:C4872              1      0 400000          7.766          7.139  1
## 2 NCIT:C4872              1 400000 1400000          9.741         13.709  2
## 3 NCIT:C4872              1 1400000 2400000          7.082         15.791  3
## 4 NCIT:C4872              1 2400000 3400000          9.000         26.829  4
## 5 NCIT:C4872              1 3400000 4400000          7.544         25.747  5
## 6 NCIT:C4872              1 4400000 5400000          6.044         25.297  6
```

Dimension of this matrix?

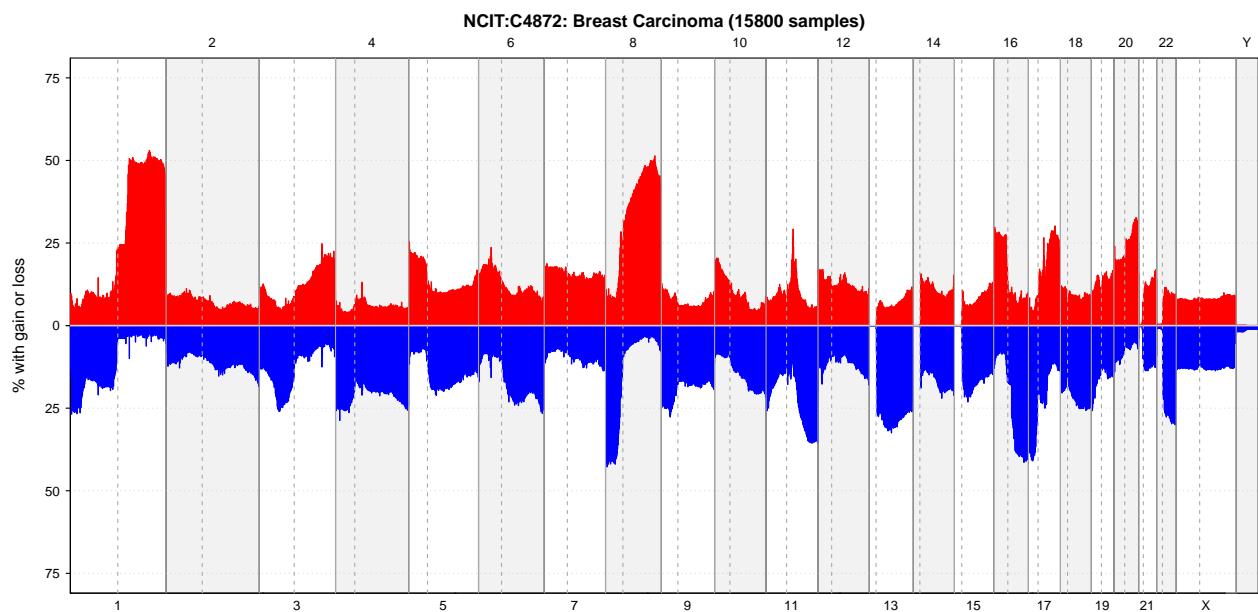
```
dim(freq$data$`NCIT:C4872`)
```

```
## [1] 3106    7
```

Step4: Visualize data

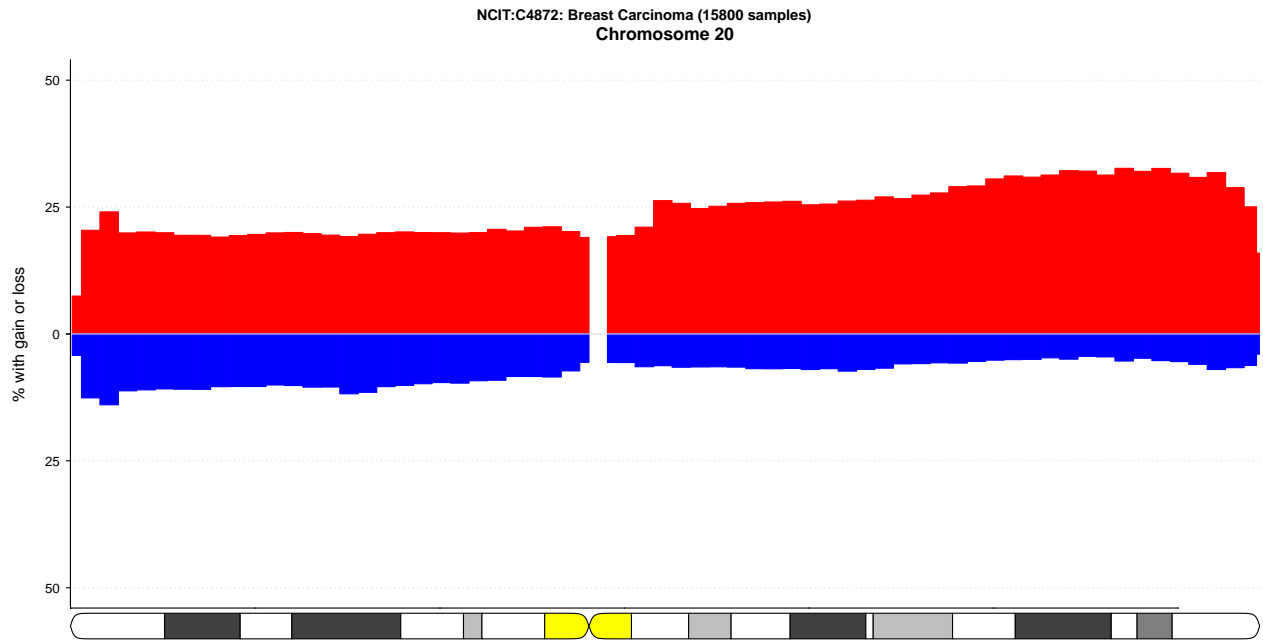
By genome

```
pgxFreqplot(freq)
```



By chromosome

```
pgxFreqplot(freq, chrom = 20)
```



Step5: Analyse the data

According the plot, we can see frequent gains on chromosome 1q, 8q, 16p, 17q, 20q and frequency losses on chromosome 1p, 8p, 11q, 13q, 15q, 16p, 22p.